UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,585	10/16/2003	Daben Liu	02-4018	5147

7590 04/16/2007

Leonard C. Suchyta
c/o Christian Andersen
Verizon Corporate Services Group Inc.
600 Hidden Ridge, HQE03H01
Irving, TX 75038

EXAMINER	
SIEDLER, DOROTHY S	

ART UNIT	PAPER NUMBER
2626	

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	04/16/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No.	Applicant(s)	
	10/685,585	LIU ET AL.	
	Examiner	Art Unit	
	Dorothy Sarah Siedler	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 October 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-31* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-31* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *16 October 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *1-20-04*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This is the initial response to the office action filled October 16, 2003. Claims 1-31 are pending and are considered below.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

> **Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.**

Claims 1-22 and 28-31 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1,10,16 and 28 fall within a judicial exception as they merely manipulate an abstract idea (mathematical algorithm) without a claimed limitation to a practical application. The claimed method is merely a series of steps to be performed on a computer, which manipulates a mathematical algorithm without any claimed limitation to a practical application.

Claims 2-9, 11-15,17-22 and 29-31 fail to resolve the deficiencies of claims 1,10,16 and 28, and therefore are rejected under similar grounds, i.e. lacking a claimed limitation to a practical application.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> **A person shall be entitled to a patent unless –**

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1,3,6,7,10-12,14,15 and 28 are rejected under 35 U.S.C. 102(b) as being anticipated by ***Liu*** ("Fast Speaker Change Detection for Broadcast News Transcription and Indexing" 1999).

As per claims 1 and 28, ***Liu*** discloses a method for classifying an audio signal containing speech information, the method comprising: receiving the audio signal (Figure 2); classifying a sound in the audio signal as a vowel class when a first phoneme-based model determines that the sound corresponds to a sound represented by a set of phonemes that define vowels (Section 3 Phone-Class Decode, paragraphs 3 and 4); classifying the sound in the audio signal as a fricative class when a second phoneme-based model determines that the sound corresponds to a sound represented by a set of phonemes that define consonants (Section 3 Phone-Class Decode, paragraphs 3 and 4, *fricatives are classified using a phoneme model: Since fricatives are by definition a specific type of consonant, it is inherent that they define consonants*); and classifying the sound in the audio signal based on at least one non-phoneme based model (Section 3 Phone-Class Decode, paragraph 3 and 4, *models are trained to classify non-speech, for example noise, music, laughter etc.*).

As per claim 10, ***Liu*** discloses a method of training audio classification models, the method comprising: receiving a training audio signal (Figure 2); receiving phoneme

classes corresponding to the training audio signal (Section 3 Phone-Class Decode, paragraphs 3 and 4,*45 context-independent HMM phone models are trained, with models for vowels, fricatives, etc. It is inherent that phoneme classes corresponding to the training audio signal are received*); training a first Hidden Markov Model (HMM), based on the training audio signal and the phoneme classes, to classify speech as belonging to a vowel class when the first HMM determines that the speech corresponds to a sound represented by a set of phonemes that define vowels (Section 3 Phone-Class Decode, paragraphs 3 and 4); and training a second HMM, based on the training audio signal and the phoneme classes, to classify speech as belonging to a fricative class when the second HMM determines that the speech corresponds to a sound represented by a set of phonemes that define consonants .

As per claim 3, ***Liu*** discloses the method of claim 1, wherein the at least one non-phoneme based model includes a model for classifying the sound in the audio signal as silence (Section 3 Phone-Class Decode, paragraph 3).

As per claims 6 and 15, ***Liu*** discloses the method of claims 1 and 10, wherein the fricative class includes phonemes that relate to fricatives and obstruents (Section 3 Phone-Class Decode, paragraphs 3 and 4, *fricatives are a specific type of obstruent, therefore it is inherent that the class includes phonemes that relate to obstrunets*).

As per claim 7, ***Liu*** discloses the method of claim 1, wherein the first and second phoneme-based models are Hidden Markov Models (Section 3 Phone-Class Decode, paragraphs 3 and 4).

As per claim 11, ***Liu*** discloses the method of claim 10, wherein the phoneme classes include information that defines word boundaries (Section 5 Experiments and Results, Word-Error-Rate (WER), *the system determines the word error rate, or word recognition accuracy. Therefore it is inherent that the phoneme classes include information on word boundaries*).

As per claim 12, ***Liu*** discloses the method of claim 11, wherein the method further comprises: receiving a sequence of transcribed words corresponding to the audio signal (Section 2 Evaluation Metrics, last paragraph, *reference transcription*); and generating the information that defines the word boundaries based on the transcribed words (Section 2 Evaluation Metrics, last paragraph, *the reference transcription is aligned with the acoustic data*).

As per claim 14, ***Liu*** discloses the method of claim 10, further comprising: training at least one model to classify the sound based on gender of a speaker of the sound (Section 3 Phone-Class Decode, paragraphs 3 and 4).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 8.8 and 31 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu***.

As per claims 8 and 31, ***Liu*** discloses the method of claims 1 and 28, but ***Liu*** does not explicitly disclose classifying the sound in the audio signal as a coughing class when the sound corresponds to a non-speech sound. However, ***Liu*** does disclose coughing as a common non-speech sound (Section 2 Evaluation Metrics, second paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to classify the sound in the audio signal as a coughing class when the sound corresponds to a non-speech sound in ***Liu***, since classification of coughing as non-speech enables the exclusion of those frames during speaker clustering for identifying speakers, as taught by ***Liu*** (Section 3 Phone-Class Decode, first paragraph).

As per claim 9, ***Liu*** discloses the method of claim 8, wherein the non-speech sound includes at least one of coughing, laughter, breath, and lip-smack (Section 2 Evaluation Metrics, second paragraph).

Claims 2,4,5,13,16-22,29 and 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Leung*** ("A Comparative Study of Signal Representations and Classification Techniques for Speech Recognition" IEEE 1993).

As per claim 16, ***Liu*** discloses an audio classification device comprising: and a decoder configured to classify portions of the audio signal as belonging to at least one of a plurality of classes, the classes including a first phoneme-based class that applies to the audio signal when a portion of the audio signal corresponds to a sound, represented by a set of phonemes that define vowels (Section 3 Phone-Class Decode, paragraphs 3 and 4), a second phoneme-based class that applies to the audio signal when a portion of the audio signal corresponds to a sound represented by a set of phonemes that define consonants (Section 3 Phone-Class Decode, paragraphs 3 and 4, *fricatives are classified using a phoneme model. Since fricatives are by definition a specific type of consonant, it is inherent that they define consonants*), and at least one non-phoneme class (Section 3 Phone-Class Decode, paragraph 3 and 4, *models are trained to classify non-speech, for example noise, music, laughter etc.*). However, ***Liu*** does not explicitly disclose a signal analysis component configured to receive an audio signal and process the audio signal by at least one of converting the audio signal to the

frequency domain and generating cepstral features for the audio signal. ***Leung*** discloses a signal analysis component configured to receive an audio signal and process the audio signal by at least one of converting the audio signal to the frequency domain and generating cepstral features for the audio signal (page 680, Abstract, *the system performs spectral and cepstral processing techniques. Therefore it must convert the signal into the frequency domain*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to process the audio signal by at least one of converting the audio signal to the frequency domain and generating cepstral features for the audio signal in ***Liu***, since spectral and cepstral feature extraction are known techniques for signal analysis, thus removing the need to spend time and resources developing a new signal analysis technique.

As per claim 2 and 19, ***Liu*** and ***Liu*** in view of ***Leung*** disclose the method of claims 1 and 16, and ***Liu*** further disclose wherein the at least one non-phoneme based model includes models for classifying the sound in the audio signal based speaker gender (Section 3 Phone-Class Decode, paragraph 3). However ***Liu*** does not disclose wherein the at least one non-phoneme based model includes models for classifying the sound in the audio signal based on bandwidth. ***Leung*** discloses an evaluation of classification techniques for speech recognition, including a comparison between telephone quality and wide-band versions of the same speech (page 680, Introduction, last paragraph). ***Leung*** discloses that the effectiveness of the classification technique may depend on

the quality of the speech signal (page 682, first paragraph), and that the telephone network inflates the phonetic classification error rate (page 682, first paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to classifying the sound in the audio signal based on bandwidth in ***Liu***, since it would enable the system to determine the optimum classifier to use based on signal characteristics, as indicated in ***Leung*** (page 682, first and second paragraph, and Figures 1 and 2, *the figure indicate the best classifier and features to use depending on the type of signal*).

As per claims 4 and 29, ***Liu*** discloses the method of claims 1 and 28,however ***Liu*** does not explicitly disclose initially converting the audio signal into a frequency domain signal. ***Leung*** discloses initially converting the audio signal into a frequency domain signal (page 680, Abstract, *the system performs spectral processing techniques. Therefore it must convert the signal into the frequency domain*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to initially convert the audio signal into a frequency domain signal in ***Liu***, sine it is a known technique for signal analysis, thus removing the need to spend time and resources developing a new signal analysis technique.

As per claims 5 and 30, ***Liu*** discloses the method of claims 1 and 28, however ***Liu*** does not explicitly disclose generating cepstral features for the audio signal. ***Leung*** discloses

generating cepstral features for the audio signal (page 680, Abstract, *the system performs spectral and cepstral processing techniques*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to generate cepstral features for the audio signal in ***Liu***, sine it is a known technique for signal analysis, thus removing the need to spend time and resources developing a new signal analysis technique.

As per claim 13, ***Liu*** discloses the method of claim 10, however ***Liu*** does not disclose training at least one model to classify the sound based on a bandwidth of the sound. ***Leung*** discloses an evaluation of classification techniques for speech recognition, including a comparison between telephone quality and wide-band versions of the same speech (page 680, Introduction, last paragraph). ***Leung*** discloses that the effectiveness of the classification technique may depend on the quality of the speech signal (page 682, first paragraph), and that the telephone network inflates the phonetic classification error rate (page 682, first paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to train at least one model to classify the sound based on a bandwidth of the sound in ***Liu***, since it would enable the system to determine the optimum classifier to use based on signal characteristics, as indicated in ***Leung*** (page 682, first and second paragraph, and Figures 1 and 2, *the figure indicate the best classifier and features to use depending on the type of signal*).

As per claim 17, ***Liu*** in view of ***Leung*** disclose the audio classification device of claim 16, and ***Liu*** further discloses wherein the second phoneme-based class includes fricative phonemes and obstruent phonemes (Section 3 Phone-Class Decode, paragraphs 3 and 4, *fricatives are a specific type of obstruent, therefore the class must include phonemes that relate to obstrunets*).

As per claim 18, ***Liu*** in view of ***Leung*** disclose the audio classification device of claim 16, and ***Liu*** further discloses wherein the first and second phoneme-based classes are determined based on Hidden Markov Models (Section 3 Phone-Class Decode, paragraph 3).

As per claim 20, ***Liu*** in view of ***Leung*** disclose the audio classification device of claim 16, and ***Liu*** further discloses wherein the decoder determines the at least one non-phoneme class using a model that classifies the portions of the audio signal as silence (Section 3 Phone-Class Decode, paragraph 3).

As per claim 21, ***Liu*** in view of ***Leung*** disclose the audio classification device of claim 16, and ***Liu*** further discloses wherein the plurality of classes additionally include: a third phoneme-based class that applies to the audio signal when a portion of the audio signal

corresponds to a non-speech sound (Section 3 Phone-Class Decode, paragraph 3 and 4, *models are trained to classify non-speech, for example noise, music, laughter etc.*).

As per claim 22, ***Liu*** in view of ***Leung*** disclose the audio classification device of claim 21, and ***Liu*** further discloses wherein the non-speech sound includes at least one of coughing, laughter, breath, and lip-smack (Section 2 Evaluation Metrics, second paragraph).

Claims 23-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Leung*** and further in view of ***Colbath*** ("Spoken Documents: Creating Searchable Archives from Continuous Audio" 2000).

As per claim 23, ***Liu*** discloses a system comprising: audio classification logic configured to classify the input audio data into at least one of a plurality of broad audio classes, the broad audio classes including a phoneme-based vowel class (Section 3 Phone-Class Decode, paragraphs 3 and 4), a phoneme-based fricative class (Section 3 Phone-Class Decode, paragraphs 3 and 4), and a non-phoneme based gender class (Section 3 Phone-Class Decode, paragraphs 3 and 4). ***Liu*** does not disclose an indexer configured to receive input audio data and generate a rich transcription from the audio data, the indexer including: a non-phoneme based bandwidth class, a speech recognition component configured to generate the rich transcription based on the broad audio classes determined by the audio classification logic, a memory system for storing the

rich transcription, and a server configured to receive requests for documents and respond to the requests by transmitting one or more of the rich transcriptions that match the requests. ***Leung*** discloses an evaluation of classification techniques for speech recognition, including a comparison between telephone quality and wide-band versions of the same speech (page 680, Introduction, last paragraph). ***Leung*** discloses that the effectiveness of the classification technique may depend on the quality of the speech signal (page 682, first paragraph), and that the telephone network inflates the phonetic classification error rate (page 682, first paragraph). In addition, ***Colbath*** discloses an indexer configured to receive input audio data and generate a rich transcription from the audio data (page 2, Component Technologies, first paragraph and page 4, System Architecture, first paragraph), a speech recognition component configured to generate the rich transcription based on the broad audio classes determined by the audio classification logic (page 2, Component Technologies, first paragraph), a memory system for storing the rich transcription, and a server configured to receive requests for documents and respond to the requests by transmitting one or more of the rich transcriptions that match the requests (page 4-5, System Architecture, server and browser).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to classifying the sound in the audio signal based on bandwidth in ***Liu***, since it would enable the system to determine the optimum classifier to use based on signal characteristics, as indicated in ***Leung*** (page 682, first and second paragraph, and

Figures 1 and 2, *the figure indicate the best classifier and features to use depending on the type of signal*).

In addition it would have been obvious to one of ordinary skill in the art at the time of the invention to have an indexer configured to receive input audio data and generate a rich transcription from the audio data, a speech recognition component configured to generate the rich transcription based on the broad audio classes determined by the audio classification logic, a memory system for storing the rich transcription, and a server configured to receive requests for documents and respond to the requests by transmitting one or more of the rich transcriptions that match the requests in ***Liu***, since it would create a system that integrates acoustic and linguistic technologies to construct a structural summary of continuous audio that is searchable by content, as indicated in ***Colbath*** (page 2, fourth paragraph).

As per claim 24, ***Liu*** in view of ***Leung*** further in view of ***Colbath*** disclose the system of claim 23, however ***Liu*** does not explicitly disclose wherein the broad audio classes further include a phoneme-based coughing class. ***Liu*** does disclose coughing as a common non-speech sound (Section 2 Evaluation Metrics, second paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the broad audio classes further include a coughing class in ***Liu***, since classification of coughing as non-speech enables the exclusion of those frames

during speaker clustering for identifying speakers, as taught by ***Liu*** (Section 3 Phone-Class Decode, first paragraph).

As per claim 25, ***Liu*** in view of ***Leung*** further in view of ***Colbath*** disclose the system of claim 24, however ***Liu*** does not explicitly disclose wherein the coughing class includes sounds relating to coughing, laughter, breath, and lip-smack. ***Liu*** does disclose coughing, laughter, breath and lip-smack as a common non-speech sounds (Section 2 Evaluation Metrics, second paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the coughing class include sounds relating to coughing, laughter, breath, and lip-smack in ***Liu***, since classification of non-speech enables the exclusion of those frames during speaker clustering for identifying speakers, as taught by ***Liu*** (Section 3 Phone-Class Decode, first paragraph).

As per claim 26, ***Liu*** in view of ***Leung*** further in view of ***Colbath*** disclose the system of claim 23, and Liu further discloses wherein the phoneme-based fricative class includes phonemes that define fricative or obstruent sounds (Section 3 Phone-Class Decode, paragraphs 3 and 4, *fricatives are a specific type of obstruent, therefore it is inherent that the class includes phonemes that relate to obstrunets*).

As per claim 27, ***Liu*** in view of ***Leung*** further in view of ***Colbath*** disclose the system of claim 23, and ***Colbath*** further discloses wherein the indexer further includes at least one of: a speaker clustering component, a speaker identification component, a name spotting component, and a topic classification component (page 2, Component Technologies).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have an indexer include at least one of: a speaker clustering component, a speaker identification component, a name spotting component, and a topic classification component in ***Liu***, since it would create a system that integrates acoustic and linguistic technologies to construct a structural summary of continuous audio that is searchable by content, as indicated in ***Colbath*** (page 2, fourth paragraph).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

- Chigier (5,638,487) discloses a speech recognition system that classifies frames of input speech
- McKiel (5,897,614) discloses a device for sibilant classification in a speech recognition system.
- Pauws (6,208,967) discloses a HMM phoneme segmentation system.
- Gupta (6,243,680) discloses a system that generates transcription from multiple utterances.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER